

Micah Sparacio
December 13, 2002

**Mental Realism:
Rejecting the Causal Closure Thesis and Expanding our Physical Ontology**

In this paper I argue for a non-reductive theory of mental causation that denies the primary tenet of contemporary theories of mind: the causal closure of the first-order physical world. I begin by briefly considering non-reductive theories of mental causation, coming to the conclusion that they all succumb to the causal exclusion problem elucidated by Jaegwon Kim. I contend that the exclusion problem highlights a fundamental error in contemporary philosophy of mind and provides us with a choice between two general and mutually exclusive options. After providing this critique, I conclude by arguing for an expanded conception of the physical, one that acknowledges the real ontological status of higher-level emergent properties and their unique causal powers.

Non-reductive Physicalism: Why it just doesn't work

Commitment to physicalism is very common in contemporary philosophy and often serves as a guiding metaphysical principle when developing theories of ontology. The commitment involves two basic principles: 1. there is some fundamental stuff (e.g. particles, fields, strings) from which everything in the universe is made and 2. this fundamental stuff interacts in particular ways such that the properties of this stuff and their relations are responsible for each and every event (i.e. change in properties) that happens in the universe. More simply: there is physical stuff and every physical event has a physical cause.

The most widely accepted theories of mental causation over the last few decades are what some describe as non-reductive physicalism. The theories are non-reductive in the sense that they retain realism about the mind (which really does have causal influence in the world). They are physicalist because they retain the causal closure thesis.

There are several problems that plague non-reductive physicalism. I will discuss what I take to be two of the more significant issues. The first of these involves the dependency of non-reductive physicalism on the causal closure thesis. I contend that this thesis has significant flaws which warrants our reconsideration of it as a metaphysical principle. The most important of these flaws lies in the difficulty we have in specifying exactly what we mean by “physical.” What exactly is a “physical” property? Do we have a set of criteria for demarcating between physical and non-physical properties (1996, Lowe, 56)? If we are doubtful about the nature of the very things that are causally closed (the physical) then why invest so much energy in maintaining this principle? It seems that the causal closure thesis lacks explanatory power precisely because it is lacking in specificity: how do we apply the principle in a useful way if we don’t even know what a physical thing is? Ultimately, the principle ends up restricting our theories of mental causation at the service of speculative, vague and ambiguous conceptions of the physical world. It is an unnecessary constraint that artificially constricts our metaphysical options.

An additional criticism of non-reductive physicalism is found in the problem of causal exclusion. Put simply, the problem is this: if physical events have physical causes (required by the causal closure thesis) then what causal work is there left for the mental. Furthermore, it can be argued that if all mental events supervene on their base physical events, and all physical events have physical causes, then the only two causal directions

that exists are from the physical to the physical and the physical to the mental. Every physical event that had subsequently been attributed to a mental cause is now seen to have a direct physical cause. Under non-reductive physicalist theories of mind, mental causation becomes dispensable. A good analogy for the exclusion problem comes from the classic example of the singing lady who breaks glass when her voice reaches a certain pitch. It is important to note that her singing involves both a semantic property and a pitch property. The semantic meaning of the phrase, though certainly a real part of the phrase, has no causal role in the breaking of the glass. The pitch property alone is causally sufficient (1996, Lowe, 74). In just the same way, if we assume that every physical event has a physical cause, though the mental accompanies the physical, it appears to be impotent.

Once non-reductive physicalism is removed as a reasonable option for explaining mental causation, the situation regarding the nature of the mental becomes much clearer. The reason for this should now be obvious: non-reductive physicalism tries to unite two incompatible theses. How can we reasonably expect to take the causal closure thesis seriously if we insist on being realists about mental causality? It turns out that either the mental falls within our conception of the physical (and its closure) or it does not. If the former, then we are left with either scientific eliminativism or epiphenomenalism. If the latter, then we are left to make sense of a world in which the mental is real. This can involve adopting a form of dualism, or, as I advocate, expanding our conception of the physical to incorporate second-order mental properties. Doing this, however, requires that we abandon the causal closure thesis, and the notion that only first-order properties

(those properties of the fundamental particles and their aggregates) deserve a place in our ontology.

After this admittedly brief analysis of non-reductive theories of the mind, we are left with what Jaegwon Kim calls “stark choices” (118). It appears that there is no middle road to travel. Either we are realists or anti-realists about the mind (119). We can’t be both. Either the mental has real causal powers or all causation is really only first-order physical causation. If we choose the former, then we are left with the task of discovering and describing the ontology of the mental. If we choose the latter, we are stuck with trying to understand how our current physical models allow for such distinct forms of causation: how are we to explain the reality of goal-directed causation by recourse to blind, purposeless relations between fundamental, first-order properties?

It is worth briefly noting that if we retain adherence to the causal closure thesis, one of our remaining options (discussed above) is epiphenomenalism, which I believe faces two daunting tasks. Epiphenomenalism asserts that while the mental has a real ontology it is causally impotent. The first problem for epiphenomenalism, then, is in calling the mental real, if it exerts no causal influence. Indeed, as Jaegwon Kim points out, one of the most reasonable criteria for something being real is the possession of causal powers (and thus properties) (119). But if the mental is causally impotent, it must not possess properties (which are individuated and defined based on causal powers). The second problematic hurdle for the epiphenomenalist is that of explaining how, if mental states have no causal influence, we are able to acquire knowledge about these states? How do we represent or refer to these states via language if there are no (downward) causal connections between the mind and our brain (Koons, 28)? Though

epiphenomenalism may be able to survive these critiques, its value as an explanatory theory is questionable.

It now seems that there are only two legitimate options when it comes to theories of mental causation. Either, we can be strict scientific eliminativists or we can acknowledge that mental properties are real things, with a real ontology, which exert a real causal influence on the world. The eliminativist properly understands and applies the causal closure thesis, taking it to its logical conclusion. The mental realist acknowledges that mental states are fundamental and irreducible with real causal powers and thus rejects the causal closure thesis. The non-reductive physicalist, on the other hand, tries to save the best of both worlds but ultimately fails. Either, the fundamental stuff and its relations are all that exist, or our conception of the physical needs to be expanded to include the irreducible emergence of fundamental, higher level physical properties and causal influences.

The Causal Influence of the Mind

I will here provide a sketch of some of the salient features that I would like to see in a satisfying account of mental causation. In doing so, I will point out reasons for preferring mental realism to eliminativism or mind-body identity. In what proceeds I have been heavily influenced by the work of E.J. Lowe and Robert Koons.

I begin my investigation into the nature of the mental by (1) affirming the real ontological status of mental states. This helps us to account for glaring dissimilarities between first-order physical causation and what we view as mental causation. In particular, it accounts for the discreteness, generality and flexibility of goal oriented mental causes (described below). However, while affirming the ontology of mental

properties, one could still defend an epiphenomenalist view by declaring that the mental is causally impotent. But epiphenomenalism, in viewing mental states as causal dead ends, simply does not satisfy our pre-theoretical intuitions. There are additional reasons for questioning the coherency of epiphenomenalism (I've described these reasons in an earlier section). It seems clear, in opposition to both eliminativism and epiphenomenalism, that (2) the mental (whatever it is we are describing when we use the word) does indeed have a unique causal influence in the world and is part of the natural causal order of events. We must, therefore, be able to (3) differentiate between mental causes and first-order physical causes. It appears that the significant distinction is to be found in (4) the teleological nature of mental causation: it is ends oriented. The contents of intentions, desires, deliberation and decisions are typically related to future states of affairs. In contrast, (5) first-order physical causation "always appears to be undirected or blind" (1997, Lowe, 10).

As indicated in (2) above, the mental is part of the natural causal order of events. It therefore seems that rather than initiating brand new chains of physical causation "out of nowhere," (6) the mental facilitates and constrains first-order physical properties towards an end or a goal. The mental can be seen as coordinating a group of mutually independent physical events, increasing the probability that a certain end state will be reached (1997, Lowe, 10). Similarly, the mental might be conceived of as a second-order physical state which, given the proper first-order state, significantly increases the probability of reaching any one of a group of causally sufficient intermediate states which thereby bring about the intended end state. Koons has recently described this as *teleological propensity enhancement* (19-21). It should be noted that this point stands in

favor of the mental realist position over the eliminativist or reductionist views for the following reason: the specificity and discreteness of a mental cause in reference to its intended effect is lost when it is identified with a physical cause or cluster of causes (i.e. neural events). The reason for this is straightforward: teleology is not present in our modern conception of physical causation.

The teleological propensity of mental causation reveals a significant feature about the role of the mental in causal chains: (7) the desired goal is a flexible generalization and therefore can often be realized via a target set of different possible physical states. When I form the intention to raise my right hand I do not have a specific physical instance in mind: neither the precise placement of my fingers nor the precise millisecond at which I raise my arm matters. The same can be seen in a game of Scrabble. When a player lays down the letter tiles, the mental constrains the general ordering of letters but is otherwise flexible regarding how each letter is situated within its box on the game board. The end state is flexible enough to allow for widely varying, yet successful physical realizations.

My discussion of (6) and (7) above provide us with a conception of the mental as being causally efficacious by means of generalized constraint. This is significant because it allows the mental a role in the natural causal order without breaking first-order causal chains or eliminating the causal efficacy of the mental. Indeed, it allows us to (8) affirm the uniqueness, and autonomy of mental causes. The role that the mental plays in causation is fundamentally different, guiding and constraining first-order chains of physical causes towards target-sets of end-states which will fulfill particular mental goals. In the same way that a train track guides a train, the mental causes guide selected first-

order chains of physical events towards a desired goal. If the track is removed, the train does not get to its destination.

The affirmation of (8) may cause us some anxiety regarding exactly what the mental is and how it is situated within the physical world. If the mental possesses new and unique causal powers (and therefore properties) that are irreducible to any first-order properties and relations, and if we affirm (1) the real ontological status of mental properties, then at first glance it appears that we might be describing a form of substance dualism: new, autonomous “stuff” coming of the blue, with new and unique properties. However, it is my contention that worries about dualism arise in this situation because of the problematic causal closure thesis. As I see it, the causal closure thesis requires that any sort of causal emergence be reducible to the fundamental level of first-order causation: nothing above and beyond. If we insist on defining physicalism circularly, by reference to the questionable causal closure thesis, my theory is certainly not a physical one. However, if we (9) expand our physical ontology to include higher-level properties and their intrinsic causal powers, then my view stands within the physicalist picture of the world.

Under my view, (10) mental properties are emergent from complex physical systems and are thus dependent on their base physical constituents. However, these constituents do not determine the mental properties that emerge. Rather, they emerge as autonomous, second-order physical states with novel, intrinsic properties which enter into the natural causal order as generalized constraints on first-order causal chains. (11) Though they are emergent and dependent, mental states are not determined by or reducible to the relations of the first-order properties of their physical base.

The view that I sketched above has several points that stand in its favor (which I will now discuss) and two primary flaws. On the positive side, I provide a way of understanding how mental states are able to fit within the causal order without the need for interaction rules between separate substances. This is accomplished by regarding mental states as emergent physical states possessing second-order intrinsic properties (with no analogue in the first-order domain) that are dependent on and realized by certain first-order structures. The mental state's causal influence lies in its unique ability to guide and constrain first-order physical chains and thereby neither initiate new causal chains "out of thin air" nor interject "in place of" first-order causes. Rather, mental causes can be seen as streambeds or channels which provide a path in which first-order physical causes proceed. This path significantly increases the probability of reaching an end-state which satisfies the mental goal.

My view also takes seriously the sharp contrast between the results of goal directed causation and the purely blind causation of first-order physical relations. First-order chains of causation proceed in an independent, chaotic, and globally-blind fashion with no end state in mind. Rarely do such independent causal chains converge on the highly specific functional patterns which are markers of mental causation (it is possible that the origin of life is one counter example, though even this is questionable). This distinction between mental and physical causes seems to be fundamentally embedded into the way we understand the world. Indeed, our very languages take note of this clear dissimilarity between mental and first-order physical causes. For example, we are quick to point out the difference between natural death and premeditated murder or natural objects and artifacts. However, we never suppose that a murder was caused outside of

the natural order: supernaturally. Nor are we confused into thinking that newly discovered pottery artifacts required some “ghost in the machine” as their cause. Since both our natural languages and common sense ontology demarcate between the results of first-order causation and the results of mental causation, we should be suspicious of attempts to blur the line. In fact, such distinctions in our ontology are essential to our very way of life (e.g. communication). Additionally, projects such as SETI’s search for advanced extra-terrestrials would not get off the ground were it not for this distinction between causal powers. In their search, SETI is seeking patterns in radio frequencies from space that will tip them off to the unmistakable (and rare) effects of mental causation (as opposed to commonly occurring natural radio frequencies). If they could not make this fundamental distinction in their search, there would be no criteria for the discovery of intelligent life: the current criterion being that only mental causation guides first-order physical causation towards a range of potential complex end-states which satisfy a specifiable goal. Certain complex physical patterns (the end states which realize a mental goal) are typically sufficient for determining a mental cause.

Finally, this theory has the distinct advantage of affirming the generalized nature of mental causation. It is clear that the mental has causal influence without having to micro-manage each particular first-order property to reach a desired goal. Rather, the causal powers of the mental lay in the ability to increase the probability of reaching a target-set range of intermediate physical states, which subsequently realize one of a range of desired end-states. This is a significant observation, for it goes to show that the mental is not constrained to particulars. The mental is flexible and general enough so that when I sign my credit card slip I need not worry about the positioning of each molecule of ink

nor whether my previous signature exactly matches the current one. As a side note, though it is outside the scope of this paper, I think that causal flexibility may be a feature of emergence in general (consider the flexibility of biological life in overcoming obstacles in the pursuit of an end state).

Two Critiques

Now that I have laid out a model of mental realism, I will critique what I see to be the two primary shortcomings (which have persistently been on the top of my mind as I've written this paper). The first problem can be summed up in a common refrain: how does it work? How does the mental emerge, and more importantly how exactly does the mental constrain chains of first-order physical causes. Sure, I've asserted that it happens, but is that satisfying? Don't we want to fill in the explanatory details? Also, have I not resurrected the traditional problem of interaction?

Though I do not have the explanatory details to fill in, and I acknowledge that my view remains mere speculation until we have such explanatory specifications, there is a fundamental difference between my view and interactionism. Whereas interactionism posits exclusive substances that somehow need to "find common ground", my view of the mental keeps it squarely within the natural causal order. The mental is nothing more than second-order physical states as described several paragraphs above. Therefore, mental states have a causal influence in the very domain from which they emerge.

The second problem with my view can be best described in the following question: what does it mean to affirm the real ontological status of the mental? Ontology, after all, is the study of the nature of things. So what is the nature of the mental? Does it take up space? Is it some kind of field? Is it non-physical? Is it merely the functioning

of the brain? Let me first say that these are the questions that philosophers will be left to debate for years to come. However, my intuition is that our difficulty in nailing down the nature of the mental involves two conflicting observations. First, we see the effects of the mind all over the place: we cannot reasonably deny that the mind has causal influence in the world. Second, the salient features of the physical world (e.g. properties like mass and space) dominate our conception of ontology: it is hard to make sense of a “real” thing if it is not physically identifiable. In dealing with this problem, some prefer to emphasize the first point (e.g. dualists) and others the second (e.g. eliminativists or mind-brain identity theorists). Some prefer to emphasize both (e.g. non-reductive physicalists), but as I have shown this strategy runs into significant problems: either the mind is part of the physical as currently conceived or it is not. My preference is to eliminate the causal closure thesis as it is currently presented in regard to the first-order properties of fundamental particles, and to expand our notion of the physical to include the second-order properties of complex systems and the irreducible causal influence of the mental. An expanded notion of the physical allows us to view the world as an integrated whole, rather than as a disjunctive world of exclusive substances in need of interaction laws. It also allows us to take the unique causal role of the mental seriously.

Implications

If my model of mental causation bears any resemblance to reality (and I hope that it does), then there are a few salient implications that are worth highlighting. First, there are multiple ways for a mind to realize its intentions in the physical world. The flexibility principle reveals that mental causes are not wholly mechanistic in the sense of causing *particular* physical states or exact routes to the desired goal. Indeed, each time I form the

intention to take a shower, eat breakfast, or (occasionally) take a walk, the content of the intention is satisfied via widely disparate chains of physical causes. If this is a general feature of mental causation then it may have significant theological implications as well.

Second, though there is a range of physical states that will realize an end state, the range is *limited*. If we consider that the mind must hone in on this target class of realizing physical states, then we are faced with what has been called *the frame problem*. How does the mind focus on the salient target class, and ignore most other possibilities. One answer to this problem might be found in the mental acts of attention and deliberation, though it seems that these acts are only involved when novel obstacles arise along the way to our desired end state. It is often the case that our mind performs this “honing” act unconsciously, or in real-time, without hesitation. We therefore have reason to assume that the mental is at least partly constituted by some sort of unconscious mechanism. However, both philosophers and computer scientists continue to be baffled as to how this problem can be resolved via computational and mechanistic methods: the frame problem remains without solution.

Finally, E.J. Lowe has asserted elsewhere that it is the “essence of intelligence to seek patterns and to attempt to impose them upon otherwise chaotic phenomena” (1996, Lowe, 69). In light of my view on mental causation, the purpose of imposing these patterns upon otherwise independent causal chains is to satisfy some desired end-state. As has been noted, the end-state cannot be identified with its physical realization. In fact, there is often a target class of different physical states, each member of which can realize an end-state that satisfies the mental goal. By noting this relation (and distinction) between the end-state and its physical realization, we may be able to consider each part

independently and potentially infer information about one from the other. This is especially the case when we only have immediate access to the instance of a physically realized end-state. We are able to infer from certain patterns in first-order physical structures to the target class of physical end-state which satisfy a particular second-order function. For example, when confronted with the physical realization of ancient artifacts, the archaeologist is typically able to make inferences from the features of the artifact to its function, purpose, or “end.”

Conclusion

I will end by providing a brief summary of the salient features of the mental as described in my paper. In doing so, I hope to clarify just where I stand on the issue and provide others with a tool for filling in the gaps that I have not addressed. To begin, I view the mental as second-order physical states that have unique intrinsic properties that are irreducible to the properties of the first-order physical states on which they depend. In order to take this position, I reject the causal closure thesis, and believe that new and autonomous causal powers can emerge from complex (first-order) physical structures. Though the mental is dependent on the complex structures from which it emerges, its causal powers are not to be identified with and are not determined by this dependence. The causal powers of the mental are situated firmly within the natural order and exert their causal influence without breaking or initiating new first-order chains of physical causes. Rather, mental causes are contentful generalizations which direct multiple, independent causal chains towards the satisfaction of an end goal by means of some range of end states.

References

Lowe, E.J. (1996), Subjects of Experience. (Cambridge: Cambridge University Press).

Lowe, E.J. (1997), "Self, Agency and Mental Causation."

URL: <http://www.stir.ac.uk/departments/arts/philosophy/cnw/webpapers/lowe1.htm>

Kim, J. (2000), Mind in a Physical World. (Cambridge, MA: MIT Press).

Koons, R. (2002), "Emergent Teleology in Psychology, Physics and Biology."

From ISCID's symposium The Teleological Origin of Biological Information.

Abstract available at: <http://www.iscid.org/symposium-2002-robertkoons.php>